

“OLYMPICS DATA VISUALIZATION”

A

Project Report

submitted

in partial fulfillment

for the award of the Degree of

Bachelor of Technology

in Department of Information Technology



Project Mentor:

Mrs. Richa Rawal
Associate Professor, IT

Submitted By :

Aashish Jain (19ESKIT001)
Anuj Sharma (19ESKIT011)
Devkinandan Sharma (19ESKIT023)
Divyansh Sharma (19ESKIT025)

**Department of Information Technology
Swami Keshvanand Institute of Technology, M & G, Jaipur
Rajasthan Technical University, Kota
Session 2022-2023**

**Swami Keshvanand Institute of Technology,
Management & Gramothan, Jaipur
Department of Information Technology**

CERTIFICATE

This is to certify that Aashish Jain, student of B.Tech(Information Technology) 8th semester has submitted their Project Report entitled "Olympics Data Visualization" under my guidance.

Mentor

Mrs. Richa Rawal

Associate Professor, IT

Signature.....

Coordinator

Mrs. Sanju Choudhary

Associate Professor, IT

Signature.....

**Swami Keshvanand Institute of Technology,
Management & Gramothan, Jaipur
Department of Information Technology**

CERTIFICATE

This is to certify that Anuj Sharma, student of B.Tech(Information Technology) 8th semester has submitted their Project Report entitled "Olympics Data Visualization" under my guidance.

Mentor

Mrs. Richa Rawal

Associate Professor, IT

Signature.....

Coordinator

Mrs. Sanju Choudhary

Associate Professor, IT

Signature.....

Swami Keshvanand Institute of Technology,

Management & Gramothan, Jaipur

Department of Information Technology

CERTIFICATE

This is to certify that Devkinandan Sharma, student of B.Tech(Information Technology) 8th semester has submitted their Project Report entitled "Olympics Data Visualization" under my guidance.

Mentor

Mrs. Richa Rawal

Associate Professor, IT

Signature.....

Coordinator

Mrs. Sanju Choudhary

Associate Professor, IT

Signature.....

Swami Keshvanand Institute of Technology,

Management & Gramothan, Jaipur

Department of Information Technology

CERTIFICATE

This is to certify that Divyansh Sharma, student of B.Tech(Information Technology) 8th semester has submitted their Project Report entitled "Olympics Data Visualization" under my guidance.

Mentor

Mrs. Richa Rawal

Associate Professor, IT

Signature.....

Coordinator

Mrs. Sanju Choudhary

Associate Professor, IT

Signature.....

DECLARATION

We hereby declare that the report of the project entitled "Olympics Data Visualization" is a record of an original work done by us at Swami Keshvanand Institute of Technology, Management and Gramothan, Jaipur under the mentorship of Mrs. Richa Rawal (Dept. of Information Technology) and coordination of Mrs. Sanju Choudhary (Dept. of Information Technology). This project report has been submitted as the proof of original work for the partial fulfillment of the requirement for the award of the degree of Bachelor of Technology (B.Tech) in the Department of Information Technology. It has not been submitted anywhere else, under any other program to the best of our knowledge and belief.

Team Members

(Aashish Jain, 19ESKIT001)

(Anuj Sharma, 19ESKIT011)

(Devkinandan Sharma, 19ESKIT023)

(Divyansh Sharma, 19ESKIT025)

Signature

Acknowledgement

A project of such a vast coverage cannot be realized without help from numerous sources and people in the organization. We take this opportunity to express our gratitude to all those who have been helping us in making this project successful.

We are highly indebted to our faculty mentor Mrs. Richa Rawal. She has been a guide, motivator source of inspiration for us to carry out the necessary proceedings for the project to be completed successfully. We also thank our project coordinator Mrs. Sanju Choudhary for her co-operation, encouragement, valuable suggestions and critical remarks that galvanized our efforts in the right direction.

We would also like to convey our sincere thanks to Prof.Dr.Anil Choudhary, HOD, Department of Information Technology, for facilitating, motivating and supporting us during each phase of development of the project. Also, we pay our sincere gratitude to all the Faculty Members of Swami Keshvanand Institute of Technology, Management and Gramothan, Jaipur and all our Colleagues for their co-operation and support.

Last but not least we would like to thank all those who have directly or indirectly helped and cooperated in accomplishing this project.

Team Members:

(Aashish Jain, 19ESKIT001)

(Anuj Sharma, 19ESKIT011)

(Devkinandan Sharma, 19ESKIT023)

(Divyansh Sharma, 19ESKIT025)

Table of Content

1	Introduction	1
1.1	Problem Statement and Objective	1
1.2	Literature Survey /Market Survey/Investigation and Analysis	1
1.3	Introduction to Project	2
1.4	Proposed Logic / Algorithm / Business Plan / Solution / Device . . .	2
1.5	Scope of the Project	2
2	Software Requirement Specification	3
2.1	Overall Description	3
2.1.1	Product Perspective	3
2.1.1.1	User Interfaces	3
2.1.1.2	Hardware Interfaces	3
2.1.1.3	Software Interfaces	4
2.1.1.4	Operations	4
2.1.1.5	Project Functions	4
2.1.1.6	User Characteristics	5
2.1.1.7	Constraints	5
2.1.1.8	Assumption and Dependencies	5
3	System Design Specification	6
3.1	System Architecture	6
3.2	High Level Design Diagrams	6
3.2.1	Use Case Diagram.	6
3.2.2	ER Diagram	7
3.2.3	Activity Diagram	8
3.2.4	Sequence Diagram	9
3.2.5	Class Diagram	10
3.2.6	Deployment Diagram	11
3.2.7	Data Flow Diagram	12

4	Methodology and Team	13
4.1	Methodology.	13
4.2	Team Members, Roles & Responsibilities.	15
5	Centering System Testing	16
5.1	Functionality Testing	16
5.2	Performance Testing	17
6	Test Execution Summary	18
7	Project Screen Shots	19
8	Project Summary and Conclusions	25
8.1	Conclusion	25
8.2	Project Summary	26
9	Future Scope	27
	References	28

List of Figures

3.1	UseCase - Olympic Data Visualization	6
3.2	ER Diagram - Olympic Data Visualization	7
3.3	Activity Diagram - Olympic Data Visualization	8
3.4	Sequence Diagram - Olympic Data Visualization	9
3.5	Class Diagram - Olympic Data Visualization	10
3.6	Deployment Diagram - Olympic Data Visualization.	11
3.7	Data Flow Diagram - Olympic Data Visualization.	12
5.1	Performace Requirements	17
5.2	Performace Testing	17
7.1	SignUp Page- Olympic Data Visualization	19
7.2	Login Page - Olympic Data Visualization.	20
7.3	Overall Leader-board	21
7.4	Participating Country	21
7.5	Overall Excels	22
7.6	Male and Female Participant.	22
7.7	Age Distribution	23
7.8	Database 1	23
7.9	Database 2	24

List of Tables

6.1 Table to test captions and labels.	18
---	----

Chapter 1

Introduction

1.1 Problem Statement and Objective

The objective of this project is to create a comprehensive and interactive data visualization system for Olympic data, with the aim of providing meaningful insights and facilitating a better understanding of the historical trends, patterns, and performances of athletes and countries participating in the Olympic Games. The project aims to address this problem by developing an intuitive and user-friendly data visualization system. The system will leverage modern visualization techniques and tools to transform raw Olympic data into visually compelling representations, including charts, graphs, maps, and interactive dashboards. By presenting the data in a visually appealing and easily interpretable manner, the system will enable users to identify trends, correlations, and outliers in the data, as well as explore specific aspects of the Olympics, such as medal distributions, athlete performances, and host countries' legacies.

1.2 Literature Survey /Market Survey/Investigation and Analysis

Performance measures for a country in the Olympics can be predicted using their past performance. By predicting their win using the maximum value scored by them in previous participation, the chance of winning gold in 2016 has been identified. If a person wins a medal in an Olympics during a year, the chance of winning a medal in the upcoming Olympics is predicted. Having sports performance data, predicting one's future performance has been done. Their performance can also be increased if they are not performing well in certain areas, and then placing them accordingly in the training program will provide considerable measure in their outcomes.

1.3 Introduction to Project

The project aims to address this problem by developing an intuitive and user-friendly data visualization system. The system will leverage modern visualization techniques and tools to transform raw Olympic data into visually compelling representations, including charts, graphs, maps, and interactive dashboards. By presenting the data in a visually appealing and easily interpretable manner, the system will enable users to identify trends, correlations, and outliers in the data, as well as explore specific aspects of the Olympics, such as medal distributions, athlete performances, and host countries' legacies.

1.4 Proposed Logic / Algorithm / Business Plan / Solution / Device

We tried to develop the system which can be used by the people of all age group. Try to develop the user interface as simple as possible.

1.5 Scope of the Project

The basic idea of visual analytics is to visually represent the information, allowing the human to directly interact with the information, to gain insight, to draw conclusions, and to ultimately make better decisions. The visual representation of the information reduces complex cognitive work needed to perform certain tasks. People may use visual analytics tools and techniques to synthesize information and derive insight from massive, dynamic, and often conflicting data by providing timely, defensible, and understandable assessments.

Chapter 2

Software Requirement Specification

2.1 Overall Description

Visual analytics tools allow consumers to understand the relationship between variables easily and make better profitable decisions. The fields of information and scientific visualization deal with visual representations of data. The main difference among the two is that scientific visualization examines potentially huge amounts of scientific data obtained from sensors, simulations or laboratory tests

2.1.1 Product Perspective

2.1.1.1 User Interfaces

For the efficient working of the User Interface, i.e. the Front End of the system, the OS must be having at least Internet Explorer 8 installed. To login to the website.

2.1.1.2 Hardware Interfaces

For the hardware requirements, the SRS specifies the logical characteristics of each interface b/w the software product and the hardware components. It specifies the hardware requirements like memory restriction, cache size, processor, RAM etc. those are required for software to run.

1. Minimum Hardware Requirements

- Hard Disk: 20GB and Above
- RAM: 512MB and Above
- Processor: Pentium III and Above

2. Referred Hardware Requirements

- HDD 80 GB
- RAM: 512 MB
- Cache: 1 MB L1
- Cache 512 KB L2

2.1.1.3 **Software Interfaces**

1. **For Hosting** - Any Windows Operations System with DOS Support and Virtual Studio for development. Primarily Windows 8 having Dream Weaver Installed with a working LAN connection to mandatory.
2. **For Using** - Any type of operating system with a Least Internet Explorer Installed and having minimum of 521 kbps working LAN compulsorily.
3. **Web Server** - Operating System (Windows)
4. **Data Base Server** - SQL Server, Operating System (Windows)
5. **Referred Software Requirements** -
 - Front End: ReactJS
 - Back End: Python, MySql

2.1.1.4 **Operations**

This software will work on all operating systems with a web browser with active internet connection.

2.1.1.5 **Project Functions**

Enlisted below are all the major functions supported by Visual Analytics on Olympic Data along with the user classes.

1. **Consumer**

- Evaluate data

- Make decision based on evaluation

2. Admin

- Provide dataset
- Update dataset
- Integrate web service

2.1.1.6 User Characteristics

Users of the system should be able to understand the relationship between the table and the charts, graphs displayed in user interface. The user of the system are the members of Olympic Committee who can analyze data and make future decisions based on the statistic available.

2.1.1.7 Constraints

- GUI is only in English.
- Login and password is used for the identification of users.
- Limited to HTTP/HTTPS or locally during development.
- This system is working for a single server.

2.1.1.8 Assumption and Dependencies

The success of this system depends on

1. Existence of an Internet service.
2. Admins (retailers) and Users (customers) must be comfortable with computers and have enough ability to work with the product.
3. Website interface must be friendly and easy-to-use.

Chapter 3

System Design Specification

3.1 System Architecture

This document presents the architecture as a series of views; use case view, logical view, process view and deployment view. There is no separate implementation view described in this document. These are views on an underlying Unified Modeling Language (UML) model developed using Rational Rose.

3.2 High Level Design Diagrams

3.2.1 Use Case Diagram

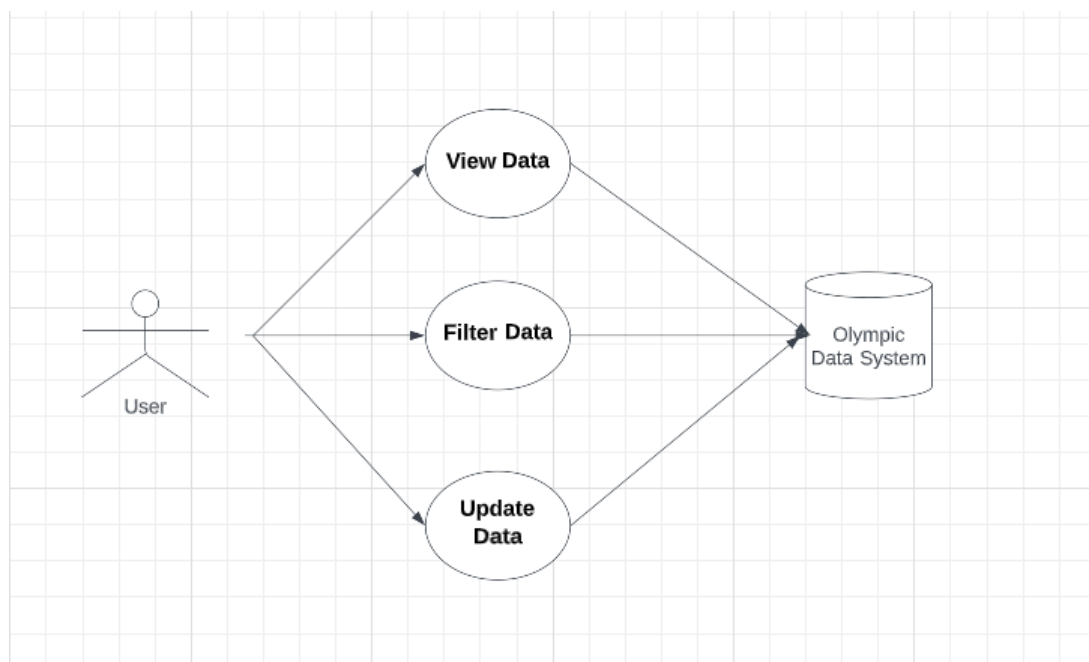


Figure 3.1: UseCase - Olympic Data Visualization

3.2.2 ER Diagram

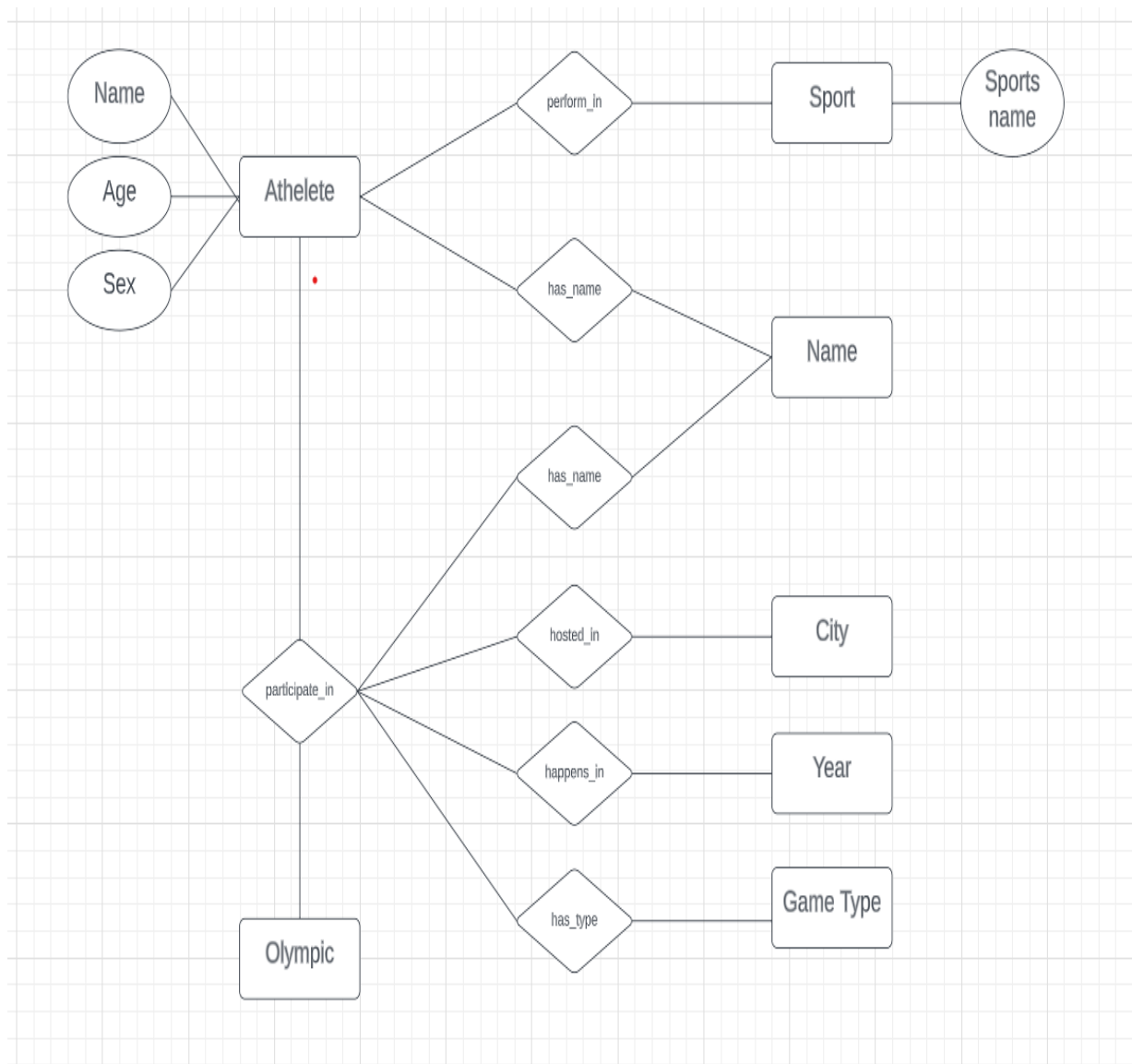


Figure 3.2: ER Diagram - Olympic Data Visualization

3.2.3 Activity Diagram

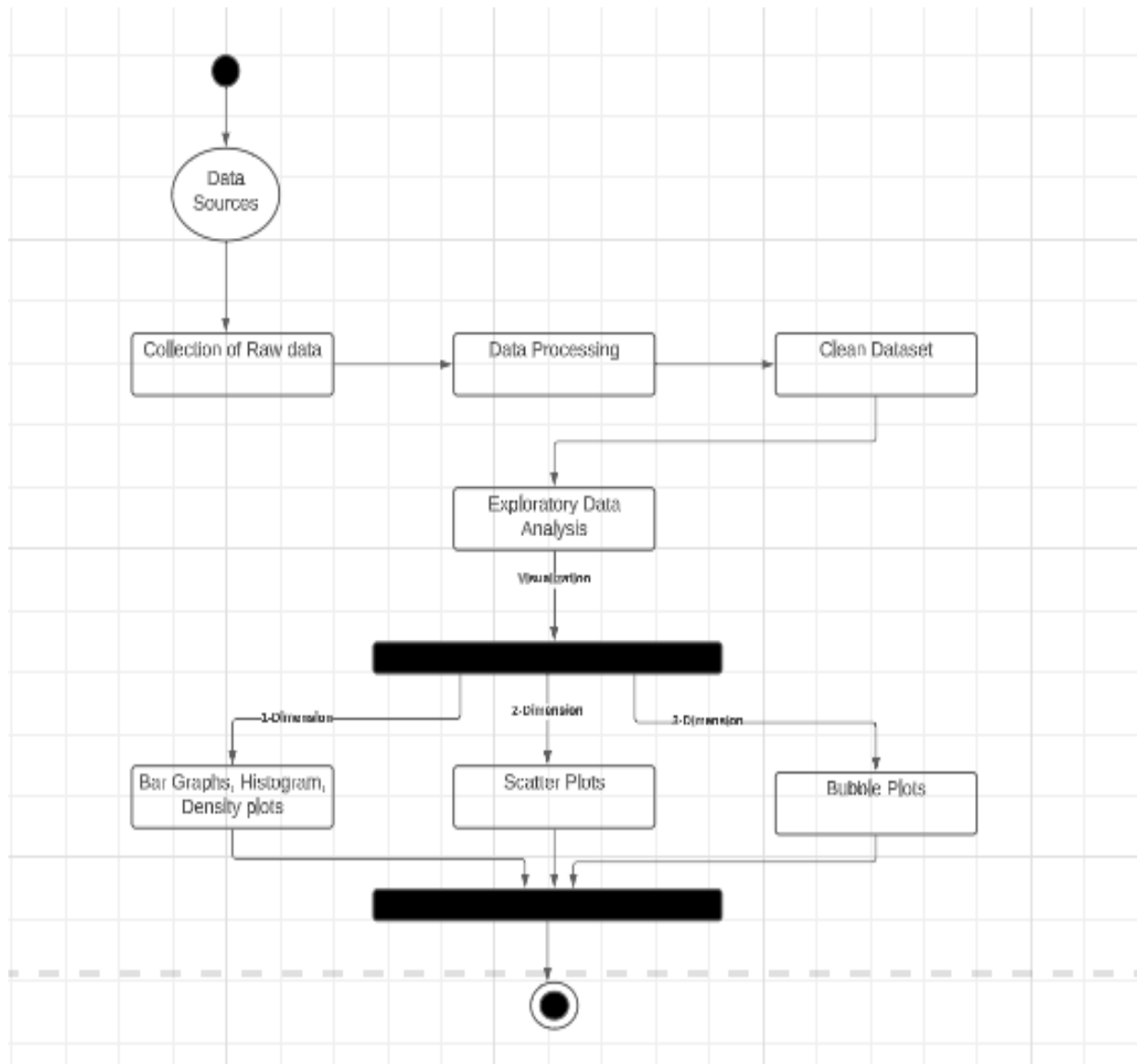


Figure 3.3: Activity Diagram - Olympic Data Visualization

3.2.4 Sequence Diagram

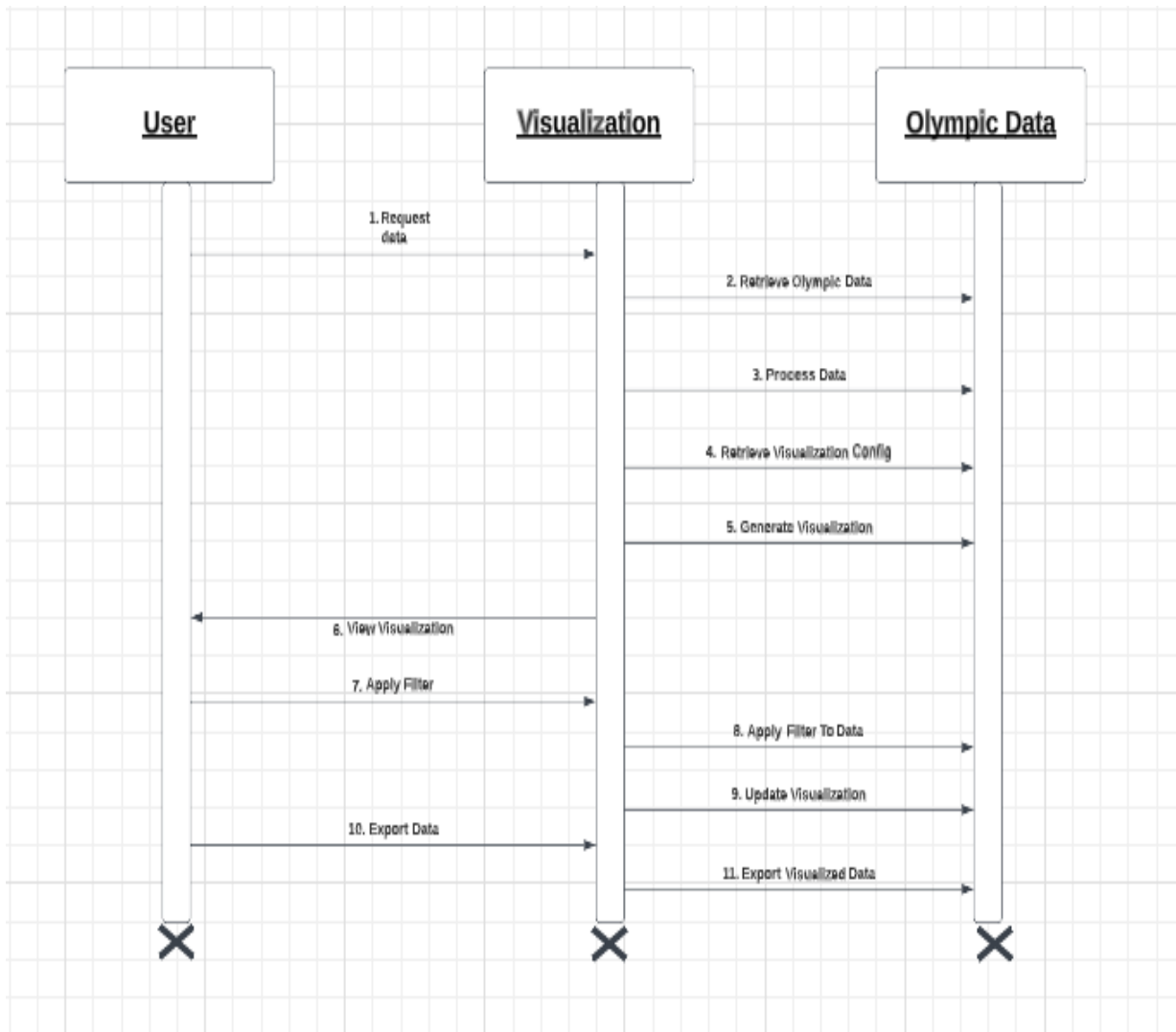


Figure 3.4: Sequence Diagram - Olympic Data Visualization

3.2.5 Class Diagram

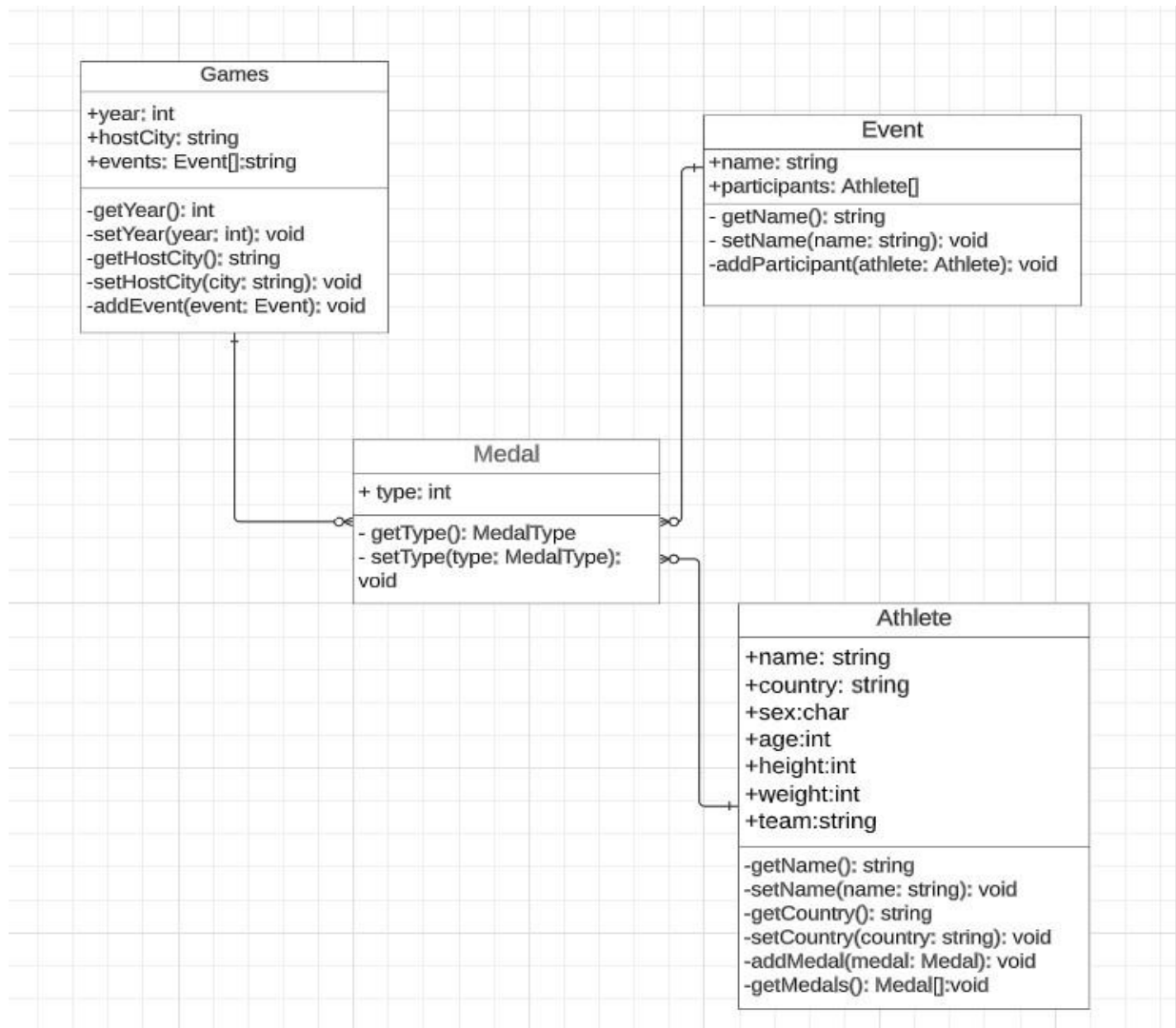


Figure 3.5: Class Diagram - Olympic Data Visualization

3.2.6 Deployment Diagram

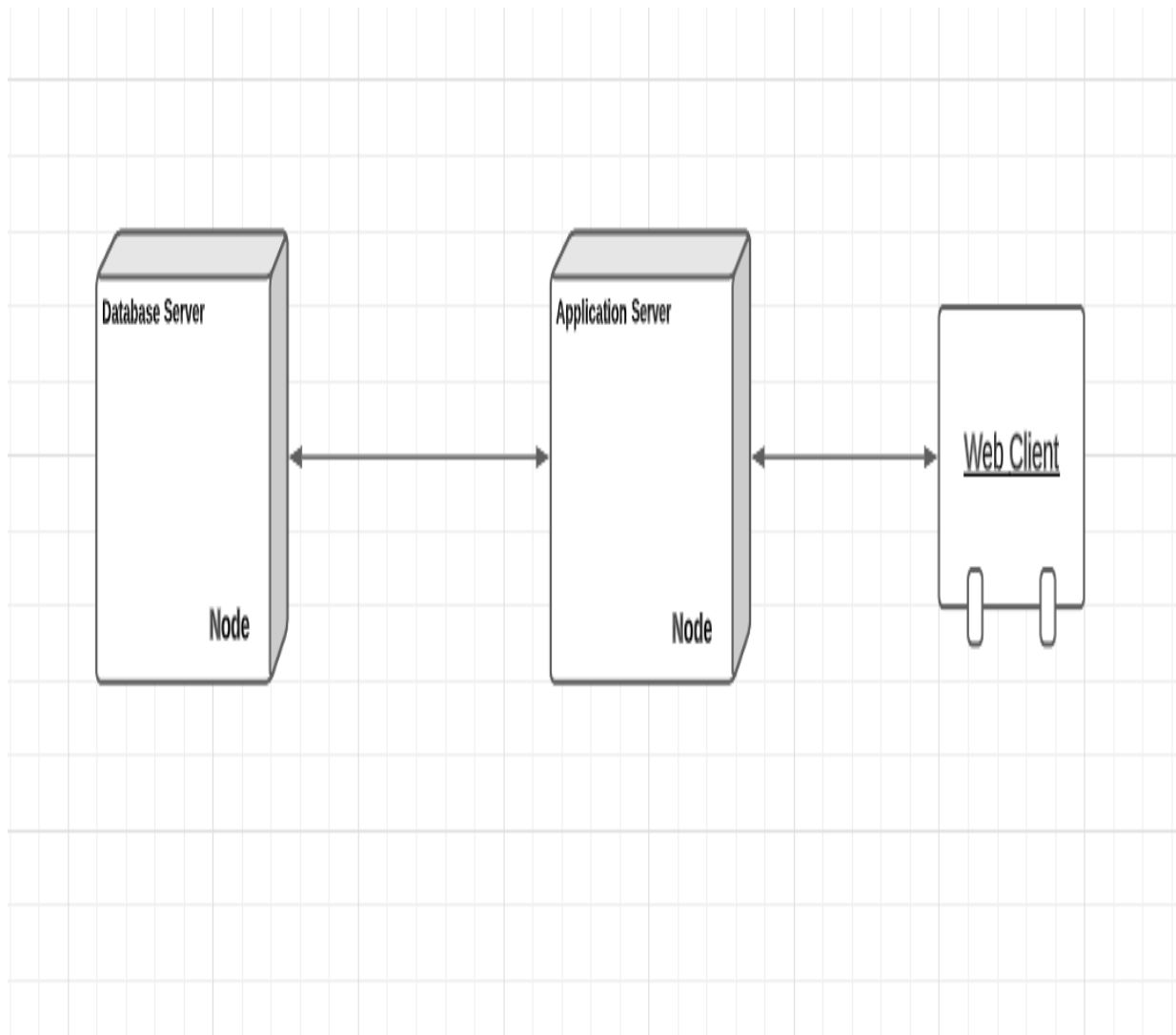


Figure 3.6: Deployment Diagram - Olympic Data Visualization

3.2.7 Data Flow Diagram

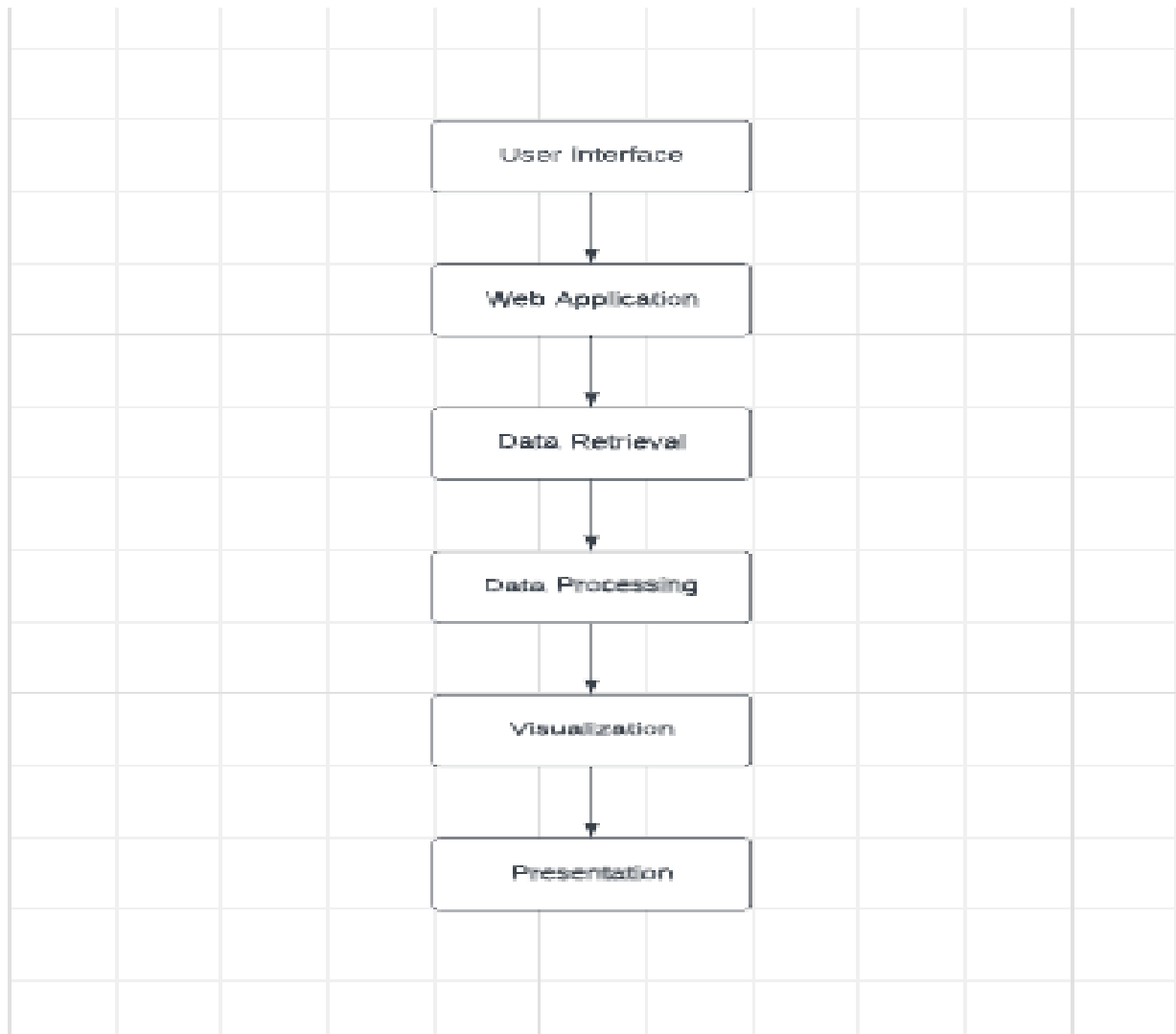


Figure 3.7: Data flow Diagram - Olympic Data Visualization

Chapter 4

Methodology and Team

4.1 Methodology

An Approach is referred to as a systematic path to reach a solution. Every problem, whether technical or non-technical, requires a proper approach so that we can get a proper path on which we have to proceed to get the required result. This Research Paper aims to analyze the vast history of Olympic Games and determine the evolution of Olympic Games over the Time. There are various factors which contribute to the evolution of the Olympics. To develop Olympics data analysis, we have followed methodology of:

1. **Data Collection** - The very first step of any type of Analysis, whether it is technical or non-technical, is Data Collection. In order to perform analysis on a certain problem, we require a large amount of Data on which we apply various techniques and algorithms to reach a particular conclusion and get our desired result. It is advised to take the data in abundance because larger the volume of data for analysis, the greater would be the accuracy in the result and also the greater would be the confidence in decision making based on these results. We have used data from various data sources for analysis on Evolution of the Olympics over the time. We have taken three datasets which provide us with a large volume and a large variety of data for Analysis. The 1st dataset consists of information about the players and their entire details like their Gender, Height, Weight, Country for which they play, Medals won (Gold, Silver and Bronze) and many more. This data can be used to analyze the performance of the particular player and can also help in the comparative study between two or more players. the 2nd dataset consists of the information of the countries which have participated in the Olympics so far and the list of the total number of medals (Gold, Silver and Bronze) won by them. This data can be used

to perform a comparative study on the performance of the countries. the 3rd dataset consists of the list of countries along with their country code which is the identification of these countries. This data can be used to find out the total number of countries which have participated in the Olympics so far.

2. Data Pre-processing The next step after collecting Data is Data Processing. Data directly obtained from a data source such as dataset is known as Raw data. We can't apply various techniques or Machine Learning Algorithms like Linear Regression, Decision Tree, SVM etc directly to the Raw Data. This Data need to be processed and converted into useful data. Data Preprocessing is the process of translating the Raw data into Useful data by conscientiously checking for errors and eliminating redundant, incomplete, or incorrect data. The Dataset consists of various fields like Age, Gender, etc which consists of some null values which produces errors in the end result which is the Visualization of data in graphical format. These null values are needed to be omitted or replaced with some valid value which solves the error and generates accurate result. We have used a technique known as Deterministic Imputation to complete this task. Deterministic Imputation is a situation where the null values (NA or NaN) are determined with the help of the other values in the same column in the dataset. For this purpose, there are various models such as Basic Numeric Imputation Model in which the null value is replaced by Mean or Median of other values of the same column of the dataset. There is another model known as Hot Deck Imputation in which the null value is replaced by similar record in the dataset, i.e., some other value in the same column. Hot Deck Imputation can be applied to both Numerical as well as the Categorical value, but only if it contains enough values in the same column.

3. Exploratory Data Analysis The next step after data pre-processing is data analysis. In this step, analysis is done on data using various Techniques like Text Analysis, Diagnostic Analysis, Exploratory Data Analysis, etc and Machine learning Algorithms like Linear Regression, Logistic Regression, SVM, Decision Tree etc to reach to a particular conclusion. As our field of Research

is visualization and comparative study of various factors which leads to the Evolution of Olympic games over the time, we are using the Exploratory Data Analysis technique to complete this task. Exploratory Data Analysis (EDA) is an approach to analyze data thoroughly and encapsulate its primary attributes basically in visual format. Exploratory Data Analysis is mainly used to see what the data represents apart from applying various algorithms. With the help of EDA, we can understand the structure and content of the dataset by various types of graphs and plots which can be drawn with the help of EDA. We can View the data in the visual format and can explain the analysis on that basis and perform a Comparative Study between different plots. There are various types of plots which are used in EDA. Some of them are mentioned below:

- Histogram
- Bar Graph
- Box Plot
- Scatter Plot and many more.

4.2 Team Members, Roles & Responsibilities

1. **Aashish Jain** - He maintained all the data for the system. Created the database for the system.
2. **Anuj Sharma** - Developed the front-end of the system. Created the front-end for the User in reactjs. Also authenticate the user by valid id and password.
3. **Devkinandan Sharma** - Responsible for developing the back-end for the system. Also fix the bugs in the UI and also help in connecting database to the back-end as well as front-end.
4. **Divyansh Sharma** - His role is to create the required document for the project. He also helped in fixing bugs in the front-end as well as back-end.

Chapter 5

Centering System Testing

The designed system has been testing through following test parameters.

5.1 Functionality Testing

In testing the functionality of the web sites the following features were tested:

1. Links

- (a) Internal Links: All internal links of the website were checked by clicking each link individually and providing the appropriate input to reach the other links within.
- (b) External Links: Till now no external links are provided on our website but for future enhancement we will provide the links to the candidate's actual profile available online and link up with the elections updates online etc.
- (c) Broken Links : Broken links are those links which so not divert the page to specific page or any page at all. By testing the links on our website, there was no link found on clicking which we did not find any page.

2. Forms

- (a) Error message for wrong input : Error messages have been displayed as and when we enter the wrong details (eg. Dates), and when we do not enter any details in the mandatory fields. For example: when we enter wrong password we get error message for acknowledging us that we have entered it wrong and when we do not enter the username and/or password we get the messages displaying the respective errors.
- (b) Optional and Mandatory fields : All the mandatory fields have been marked with a red asterisk (*) and apart from that there is a display of error messages when we do not enter the mandatory fields. For example: As the first

name is a compulsory field in all our forms so when we do not enter that in our form and submit the form we get an error message asking for us to enter details in that particular field.

3. Database Testing is done on the database connectivity.

5.2 Performance Testing

Metric	Value	Descriptions
Average Page Load Time	< 4 sec	
Max Page Load Time	< 60 sec	
Minimum Throughput	per scenario	
Max Number of Registered Users	30k	
Max Concurrent Connection	2000	
Max Records in the DB	300k	e.g. 10000 invoices, 3mln customers, etc...
Max DB Size	11GB	

Figure 5.1: Performace Requirements -

Scenario (Thread Group)	Expected Normal Throughput in peaks [scenarios/min] (100%, Load Test)	Number of concurrent threads (100%)	Throughout in JMeter plan [scenarios/min] for Load Test (100%)	Number of concurrent threads (200%)	Throughout in JMeter plan [scenarios/min] for Stress Test (200%)
Anonymous Views a front page	120	10	140	20	180
Anonymous Browse Course catalogue and an outline	60	10	50	20	65
Training Coordinator Edits a node	2	1	4	2	8

Figure 5.2: Performace Testing -

Chapter 6

Test Execution Summary

Execution Test Summary Report is an overall view of Testing Process from start to end. Test Plan comes at the starting of project while Test Summary Report comes at the end of the testing process. This report is given to the client for his understanding purpose. The Test Summary Report contents are :

1. Test Case ID generated
2. Total number of resources consumed
3. Passed Test Cases
4. Failed Test Cases
5. Status of Test Cases

S.No	Test Case Id	Test Case Description	Test Case Status	No. of Resources Consumed
1	6	87837	787	
2	7	78	5415	
3	545	778	7507	
4	545	18744	7560	
5	88	788	6344	

6.1 Table to test captions and labels

Chapter 7

Project Screen Shots

1. SignUp

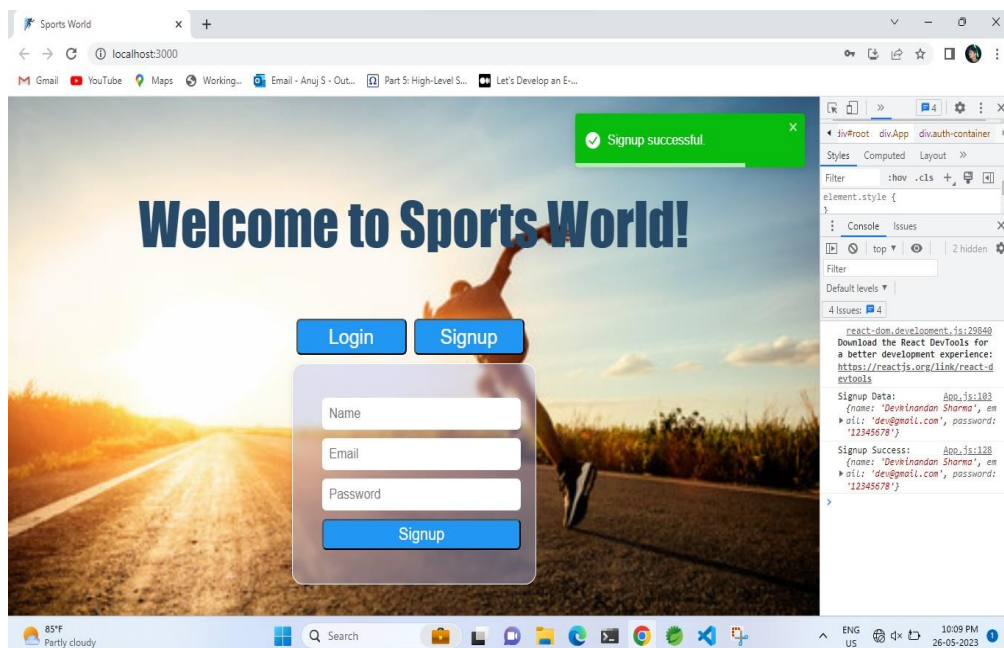


Figure 7.1: SignUp Page- Olympic Data Visualization

2. Login

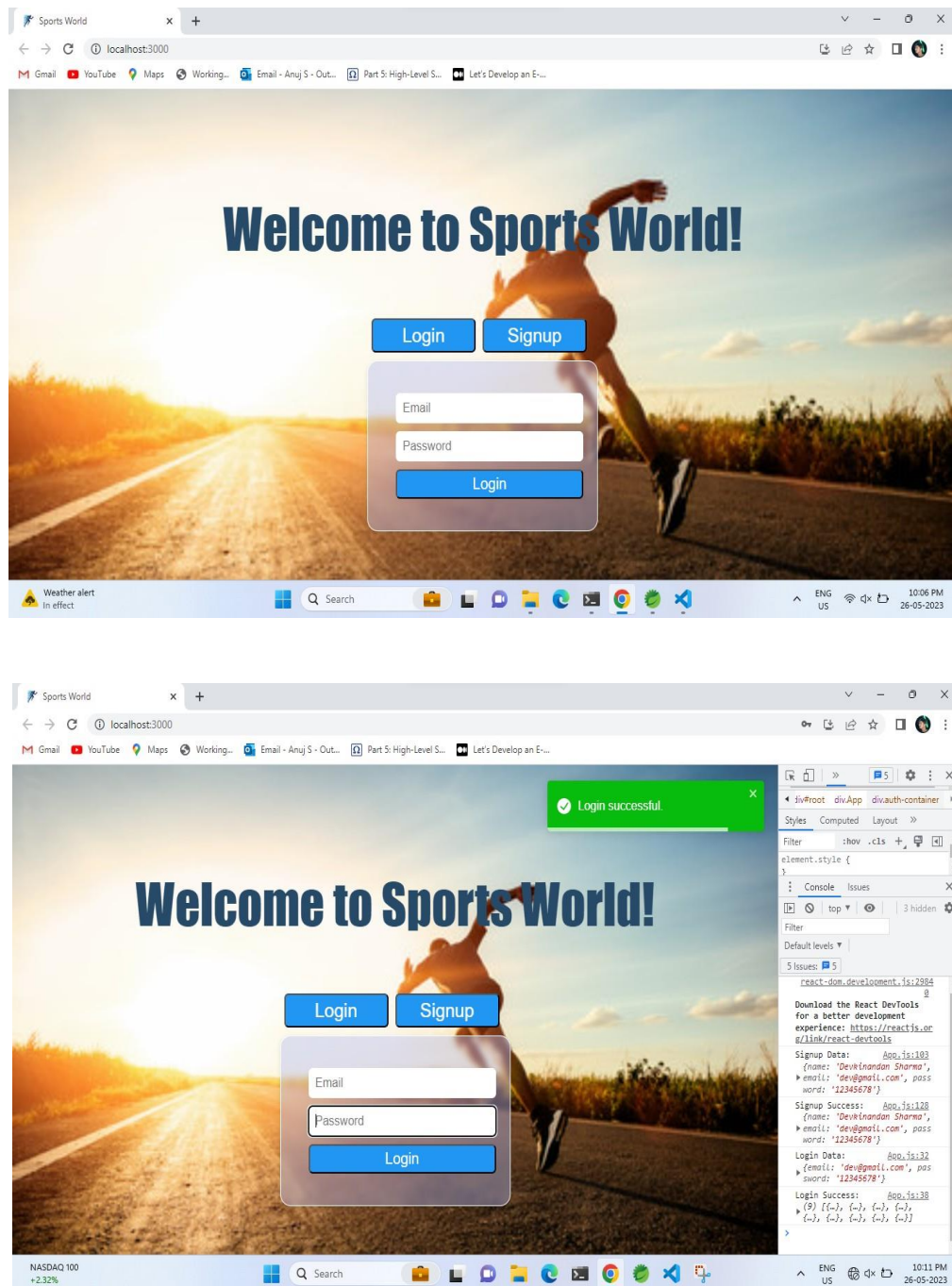


Figure 7.2: Login Page - Olympic Data Visualization

3. Overall Leader-board

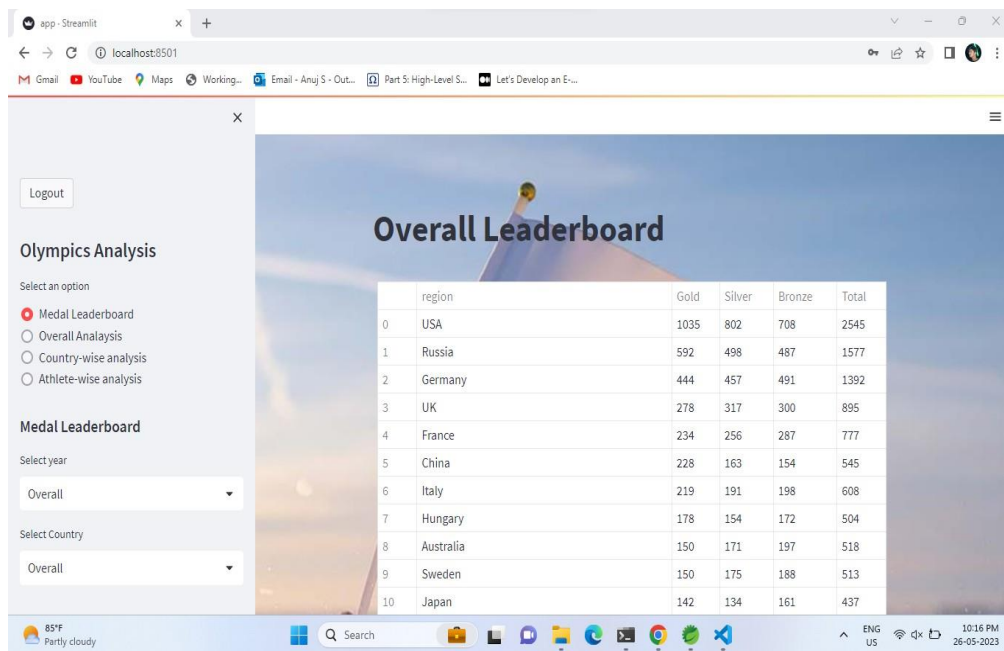


Figure 7.3: Overall Leader-board

4. Participating Country



Figure 7.4: Participating Country

5. Overall Excel

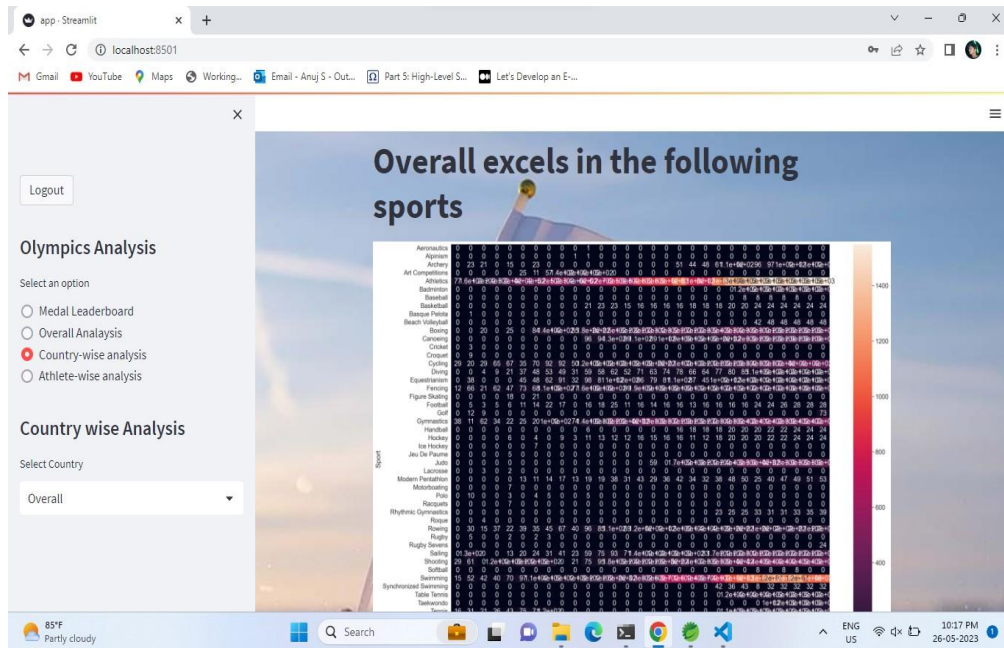


Figure 7.5: Overall Excels

6. Male and female participant

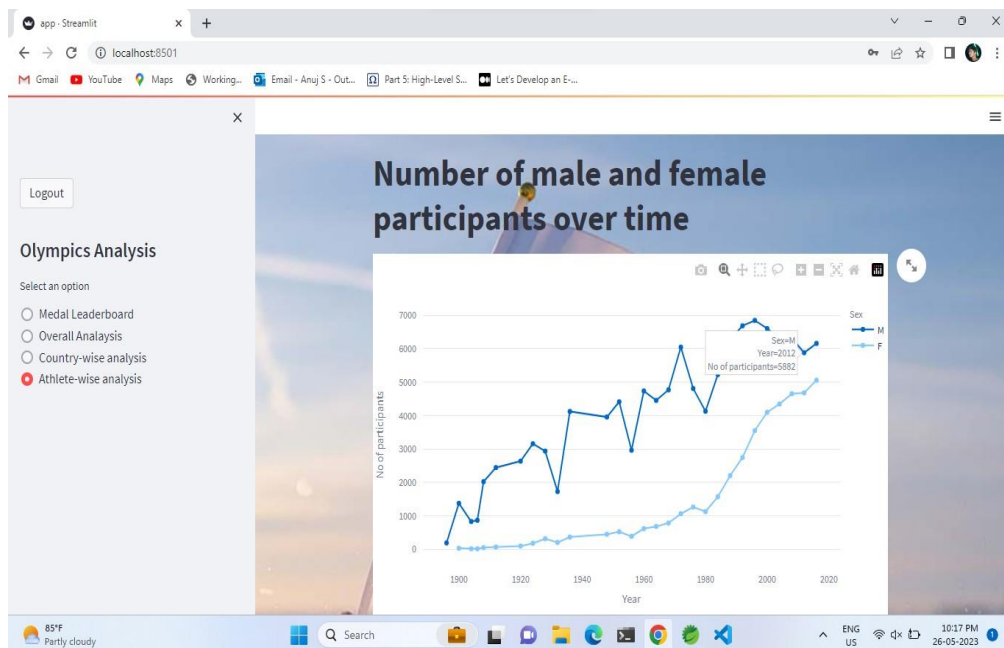


Figure 7.6: Male and Female Participant

7. Analysis based on age

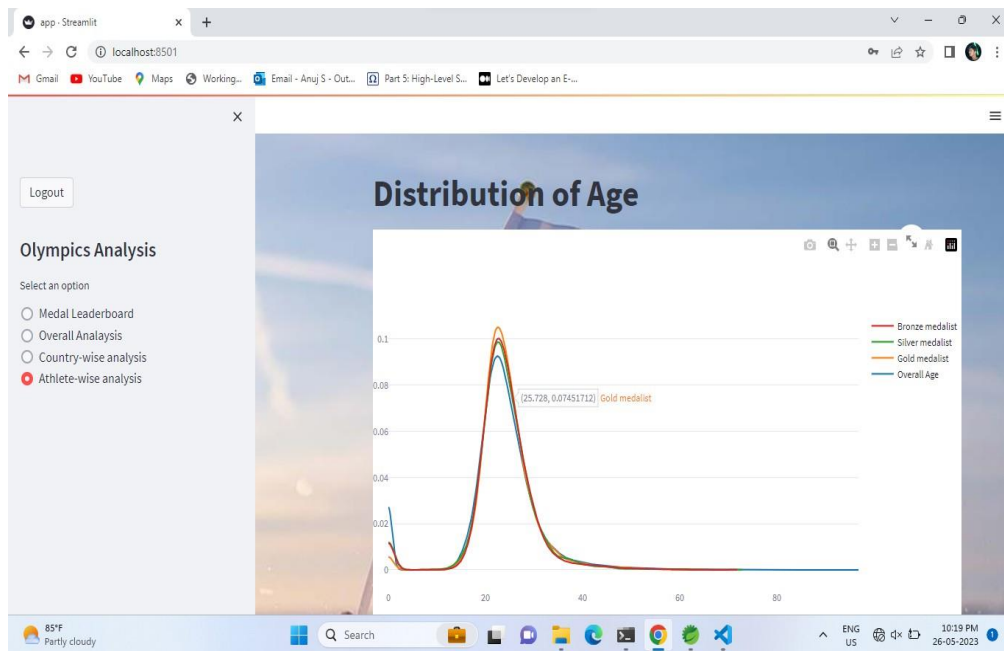


Figure 7.7: Age Distribution

8. Database of Athlete events

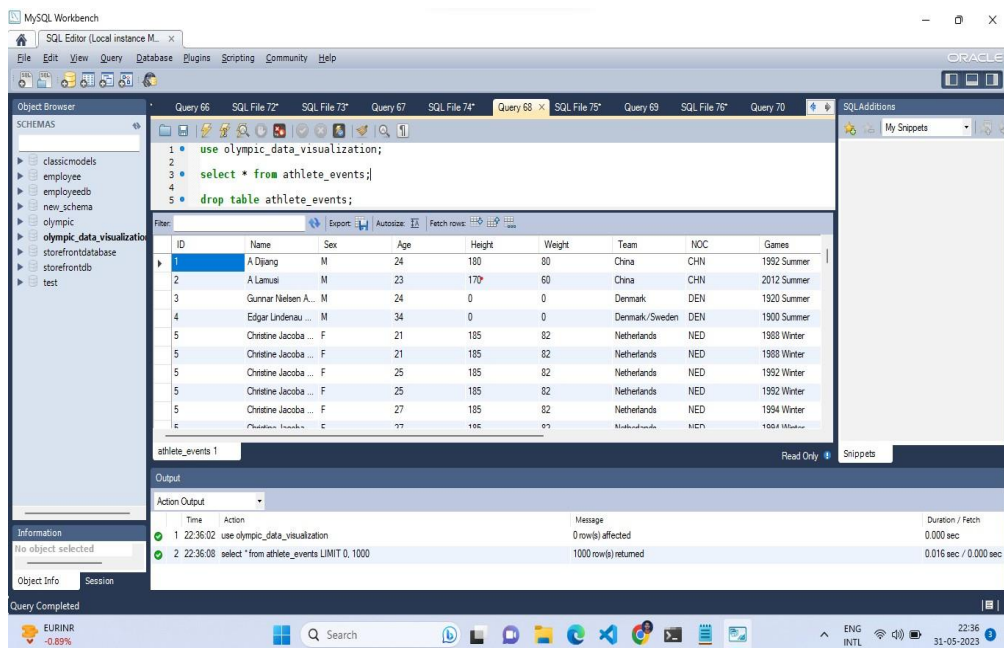


Figure 7.8: Database 1

9. Database of Noc Regions

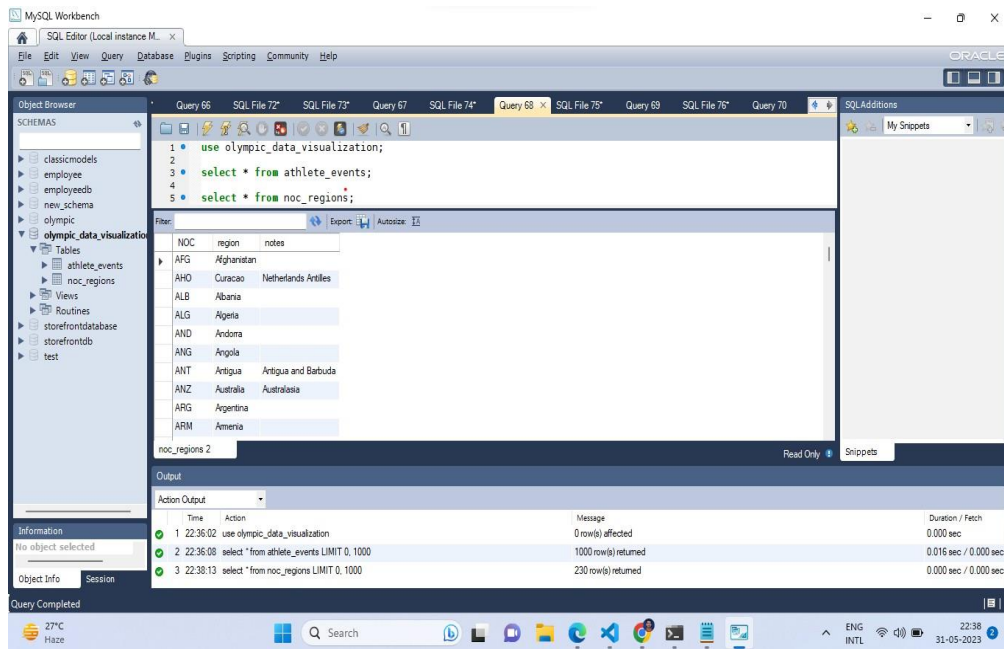


Figure 7.9: Database 2

Chapter 8

Project Summary and Conclusions

8.1 Conclusion

The main objective of this study was to analyze and visualize the various factors which have contributed to the Evolution of the Olympic Games over the years. This type of analysis is very helpful as this type of analysis can be performed by any Country or Player which can help them in analyzing their performance so that they can improve their performance by changing their strategies. We have used a technique named Exploratory Data Analysis which enables you to encapsulate the primary factors of a dataset into a visual format. We selected Python language to implement our work because It is one of the best languages suitable for Data Analysis and is the platform where we have performed this Analysis. As a result of the Analysis, we can conclude that It is true that Olympic Games have evolved considerably over time from the 1896 Olympic Games till the 2016 Rio Olympics. Various factors provide valid evidence that the Olympics have changed a lot. some of these factors are the launch of the Winter Olympic Games apart from the Summer Olympic Games in 1924, an increase in the number of participating countries in both Summer and Winter Olympics, the Average age of players in the Olympic Games, the increase in the participation of the females in both Summer and Winter Olympics over the time, Total number of medals won by various participating countries over the years, Average height and the weight of Players who contributes to victory of Games in the event. Apart from these, there are many more factors that depict the Evolution of the Olympic Games over time. Visualization of these factors has been done to explain and validate the Analysis in various Graphical formats like a Line graph, Scatter Plots, Bar, Graphs, Dist Plots, etc.

8.2 Project Summary

The proposed project is to create a comprehensive and interactive data visualization system for Olympic data, with the aim of providing meaningful insights and facilitating a better understanding of the historical trends, patterns, and performances of athletes and countries participating in the Olympic Games.

Despite the wealth of data available from various Olympic Games editions, accessing, analyzing, and comprehending the information in a coherent and visually appealing manner can be challenging. Traditional methods of data representation, such as tables and spreadsheets, often fail to capture the complexity and dynamics of Olympic data. This lack of effective data visualization hinders researchers, sports enthusiasts, and policymakers from gaining deeper insights into the Olympics' evolution and its impact on participating nations.

Chapter 9

Future Scope

We all know that any Analysis is not perfect and it consists of some limitations which define the Future scope of the Research Work. This project work also contains some limitations which we are considering as the Future Scope of the Project. We can also describe the data in other formats like Geographical format where we can depict the countries on the World map. We can also apply various Machine Learning Algorithms to the data set after Analysis and can create a Predictive Model which can predict the statistics of the Future Olympic Games.

References

- [1] Wikipedia contributors: [https://en.m.wikipedia.org/wiki/Olympic Games](https://en.m.wikipedia.org/wiki/Olympic_Games), last accessed 2020/11/02.
- [2] Dey S K, Rahman M M, Siddiqi U R and Howlader A 2020 Analyzing the epidemiological outbreak of COVID-19: A visual exploratory data analysis approach J. Med. Virol. 92 632–8
- [3] Bondu R, Cloutier V, Rosa E and Roy M 2020 An exploratory data analysis approach for assessing the sources and distribution of naturally occurring contaminants (F, Ba, Mn, As) in groundwater from southern Quebec (Canada) Appl. Geochem. 114 104500
- [4] Cutait, M.: Management performance of the Rio 2016 Summer Olympic Games. Research Paper submitted and approved to obtain the Master's degree in Sports Administration at AISTS in Lausanne, Switzerland.
- [5] Moreno A, Moragas M and Paningua R 1999 The evolution of volunteers at the Olympic Games Proceedings of Symposium on Volunteers (Lausanne, Switzerland: Global Society and the Olympic Movement) pp 1–18
- [6] Github link: github.com/anujsharmas2001/Olympic-Data-Visualization